Renewable and Sustainable Energy Reviews 23 (2013) 272–288

Contents lists available at SciVerse ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

State of the art in building modelling and energy performances prediction: A review

Aurélie Foucquier[a,*], Sylvain Robert[b], Frédéric Suard[a], Louis Stéphan[c], Arnaud Jay[c]

[a] CEA, LIST, Laboratoire Information, Modèles et Apprentissage, Gif-sur-Yvette Cedex, France
[b] CEA, LIST, Laboratoire d'Outils pour l'Analyse de Données, Gif-sur-Yvette Cedex, France
[c] CEA-INES, LITEN, Laboratoire Energétique du Bâtiment, Le Bourget Du Lac, France

ARTICLE INFO

Article history:
Received 9 November 2012
Received in revised form
26 February 2013
Accepted 2 March 2013
Available online 27 March 2013

Keywords:
Building modelling
Energy consumption
Energy performance
Building thermal models
Machine learning
Building prediction model

ABSTRACT

In the European Union, the building sector is one of the largest energy consumer with about 40% of the final energy consumption. Reducing consumption is also a sociological, technological and scientific matter. New methods have to be devised in order to support building professionals in their effort to optimize designs and to enhance energy performances. Indeed, the research field related to building modelling and energy performances prediction is very productive, involving various scientific domains. Among them, one can distinguish physics-related fields, focusing on the resolution of equations simulating building thermal behaviour and mathematics-related ones, consisting in the implementation of prediction model thanks to machine learning techniques. This paper proposes a detailed review and discussion of these works. First, the approaches based on physical ("white box") models are reviewed according three-category classification. Then, we present the main machine learning ("black box") tools used for prediction of energy consumption, heating/cooling demand, indoor temperature. Eventually, a third approach called hybrid ("grey box") method is introduced, which uses both physical and statistical techniques. The paper covers a wide range of research works, giving the base principles of each technique and numerous illustrative examples.

Contents

* Corresponding author. Tel.: +33 1 69 08 30 52.
E-mail address: aurelie.foucquier@gmail.com (A. Foucquier).

1364-0321/$ - see front matter
http://dx.doi.org/10.1016/j.rser.2013.03.004

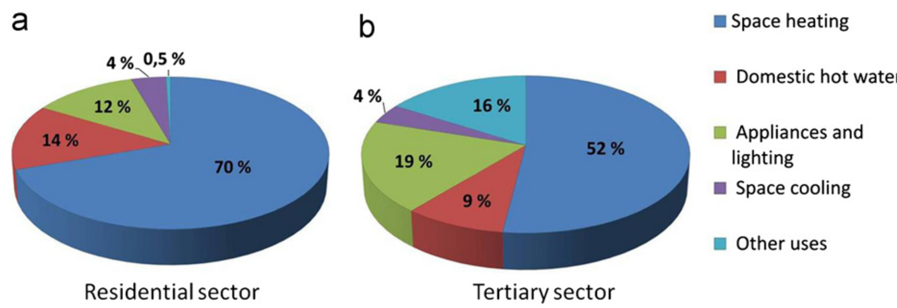

Fig. 1. (a) Scheme of the energy uses distribution in buildings in residential and tertiary sector in 2001 [3,4].

1. Introduction

The building sector in the European Union is considered as the largest consumer of energy with using up to 40% of the final energy consumption [1]. More specifically, residential uses represent about 60% of total energy consumption of the building sector [2,3]. To evaluate the energy performance of both residential and tertiary buildings, many parameters are required: thermal characteristics of the building, ventilation, passive solar system, indoor/outdoor climatic conditions and energy end-uses [2]. Considering these influencing factors, the average energy consumption in European Union raises to about 200 kW h/m^2/year, distributed as shown in Fig. 1 [3,4].

Thereby, it seems obvious to make significant efforts in terms of energy savings in building sector. For instance, the European Union established specific actions by introducing the EPBD (Energy Performance of Building Directive) dedicated to the building environmental issue [1]. This directive suggests to each EU states to target their own objectives. As a consequence, different projects of passive building emerged in Germany with PassivHaus, in Switzerland with Minergie and in France with Effinergie [5,6].

From a practical and scientific point of view, various solutions have been proposed both to increase the energy efficiency and to reduce greenhouse effects:

- An awareness campaign with the occupants on the environmental issue is necessary to reduce end-use energy consumption [7]. Simple actions could decrease significantly the energy consumption as changing the space heating behaviour, unplugging the computer or mobile charger and unused devices, configuring the computer to hibernate after a given time of inactivity, avoiding waste of hot water and many other actions [8].
- A second solution consists in the design of new dwellings or the refurbishment of existing dwellings with bringing energy-efficient improvements in agreement with the regulations given above. For instance, one way is to favour the exterior insulation and to replace simple glazing windows by double or triple glazing depending on the exposure of the room. However, the choice of energy-efficiency improvements is not obvious and the risk is to produce opposite effects.
- A third solution is to optimize the use of energetic systems as heating or cooling load. Indeed, new technologies give the

possibility to improve significantly the energy efficiency. The integration of renewable energies in these systems is also quite efficient. For instance, Badescu and Sicre [9,10] evaluated the performance of the solar energy on a passive house in Germany and showed the possibility to reduce the heating demand to 5–6 kW h/m^2/year.

- In addition to the two last proposals, a fourth solution is to use control and monitoring systems allowing controlled blackouts during specific moment of the day. Many authors have already proved the efficiency of such systems on the energy performance of the building [11–13]. For example, very recently, [14] have published a work dealing with a model-predictive control of the HVAC systems able to control the indoor temperature of a room of a computer laboratory in the University of Berkeley. Previously, Mossolly et al. [15] compared several control strategies in order to increase the energy performance in an academic building in Beirut, Lebanon. By determining the optimal control strategy, they recorded energy savings up to 30% during the summer. Moreover, these examples showed the ability to deal with very large scale systems.

The design of building integrating all these efficiency measurements is usually "tested" and validated via software taking into account these specific aspects. The aim is to predict the improvements that could be made considering different designed management. So, scientists and engineers frequently resort to various and numerous simulation techniques. Depending on the use cases, several approaches are available: some of them based on the thermal knowledge and physical equations of the building and others based on the data collected inside the building.

We propose to give an overview of these existing methods. In Section 2, we will introduce the physical techniques called "white box" approaches used to model the thermal behaviour of a building. This kind of approach is used for several applications at different scales. For example, the white box scheme allows one to evaluate the indoor temperature in a building for different time (year, month, day or hour) and spatial (the entire building, a room, a cell of a room) scales. Then, in Section 3, we will present the statistical or machine learning formulations called "black box" approaches mainly used in the aim to deduce a prediction model from a relevant database (for example, to forecast energy consumption or heating/cooling load in a given building). Finally, in Section 4, we will introduce solutions to couple the white and black box techniques to implement hybrid approaches also called "grey box" approaches.

Some of these techniques have already been referenced by Zhao and Magoulès in their review article [16]. This is true e.g. for artificial neural network and support vector machines: therefore, we will not be as exhaustive as they were in those points and invite the reader to refer to this previous review.

2. Physical models: building thermal behaviour modelling

Physical models are used to model the thermal behaviour in different varieties of buildings with their own specific needs: dwelling, office, hospital, school, firms, etc. Some of them include models of space heating [17,18], natural ventilation [19,20], air conditioning system [21], passive solar [22], photovoltaic panel [23,24], hygrothermal effects [25,26], financial issue [27], occupants behaviour [28–30], climate environment [31], etc. The physical techniques are based on the solving of equations describing the physical behaviour of the heat transfer.

These equations can be written via the energy conservation law as follows:

$$\Phi_{int} + \Phi_{source} = \Phi_{out} + \Phi_{stock} \tag{1}$$

Φ_{int} is the heat flux entering the system, Φ_{source} the heat flux of an eventual heat source, Φ_{out} the heat flux leaving the system and Φ_{stock} the heat flux stored. The principal in- and out-coming fluxes taking place in the heat transfer are the conduction through walls, the convection, the longwave and shortwave radiation and the ventilation.

To solve such physical problems, a large number of numerical software are available. Many authors proposed benchmarks to compare these software [32–36]. For this reason, we will not develop here a software comparison. Theoretically, each building software is able to include each of the mechanisms given above. They give the choice to users to select the mechanisms and the associated equations occurring in the system. But, Woloszyn [35] and Crawley [34] showed that many software are badly adapted to take into account moisture influences and generally, the effects of the latent heat are neglected.

Three main thermal building models are currently used: the multizone, zonal and CFD (Computational Fluid Dynamics) methods. We cannot say that one of these physical formulations is particularly better than another. Each of them has its own application and by this fact, the choice of the physical method depends essentially on the problem. It is precisely what we will discuss in this section. In the following, we will detail and give some examples for each of these methods.

Each sections are built in the same way with a first part describing the principle of the approach, a second part with the advantages and the application field, a third part with the limitations of the method and a last part with some examples.

2.1. The CFD approach

2.1.1. Principle of the CFD approach

The most complete approach in the thermal building simulation is the CFD (Computational Fluid Dynamics) method. This is a microscopic approach of the thermal transfer modelling allowing to detail the flow field. It is based on the decomposition of each building zone in a large number of control volumes with homogeneous or heterogeneous global mesh. Therefore, the CFD technique is recognized as a three-dimensional approach.

Software using the CFD model are essentially based on the resolution of the Navier–Stokes equation. Given that it is not the main topic of this review, we will not give more details on the CFD equations. A huge number of CFD software are available such as FLUENT [37], COMSOL Multiphysics [38], MIT-CFD, PHOENICS-CFD [39], etc. Their application fields are very large and not always specific to building simulation. Indeed, they can be applied to every systems considering a detailed flow description.

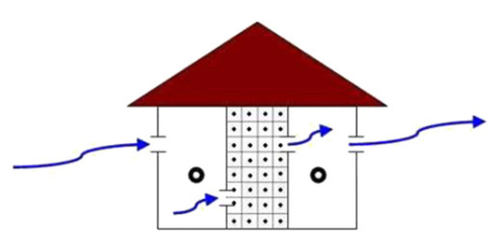

Fig. 2. Schematic representation of a problem solved with the zonal method (courtesy of Maxime Trocmé) [145].

2.1.2. Advantages and application field of the CFD

The CFD method is mainly employed for its ability to produce a detailed description of the different flows inside buildings (airflow, pollutant flow, etc.). Consequently, the CFD is very well-adapted to the study of the particle transport as pollutant particles. Moreover, as we mentioned before, the volume is divided into several discrete control volumes. Thus, it allows one to study very complex geometries of the building by minimizing locally the mesh of some specific parts.

2.1.3. Limitations of the CFD method

The main disadvantage of the CFD approach resides in its huge computation time [40], due to the fact that a complete detailed 3D-description of the building with a very fine mesh is absolutely required. Consequently, the smaller the mesh, the larger the computation time. However, given that the air velocity in at least 75% of the building is less than 0.5 m/s, it is not always necessary to apply the CFD technique in the entire building but just to specific constituents of the building as HVAC (Heating, Ventilation and Air Conditioning) equipment or appliances. Thus, it allows one to reduce considerably the computation time. For this reason, the CFD is frequently coupled with less time-consuming thermal building simulation techniques as those that will be introduced in the following subsections or else statistical techniques as those that will be presented in the second part of this paper. For example, Tan and Glicksman [40] compared the full CFD simulation and the coupling between the CFD and another building simulation method for modelling the natural ventilation across large openings or atrium. They showed that the full CFD simulation would take more than 10 hours, whereas the coupled method needs less than one hour. In the same way, Qin and Zhou [41] coupled a machine learning technique with a method coupling the CFD and a building energy model to predict the thermal dynamic behaviour in a large volume room as an atrium.

Moreover, the CFD method is quite limited by the complexity of the model implementation. Indeed, it is quite difficult to use it without previous knowledge on fluid dynamics and software. Furthermore, the CFD is also largely limited when it comes to model of the turbulence.

2.1.4. Applications reviews

Zhai et al. [42] coupled a building simulation software called EnergyPlus [43] and the CFD software MIT-CFD to predict the cooling or heating demand both in an office and in an auto racing complex. The authors used EnergyPlus to determine the cooling or heating demand and MIT-CFD to find the airflow and temperature distribution in the zone volume. At each time step, EnergyPlus passed the information to the CFD program that used them as boundary conditions. Then, the CFD program deduced the distribution of the air temperature in the thermal boundary layer and the convective heat transfer coefficients into the office. Finally, these outputs are injected in EnergyPlus as inputs to improve the accuracy of the heating load prediction.

Other authors chose the same strategy. For example, Wang and Wong [44] used a building simulation software ESP-r [45] and FLUENT (a flow software using the finite volume method) [37] to simulate the natural ventilation in residential buildings. The ESP-r simulation contained the geometrical information, the construction thermal properties and the airflow network for the whole building. The place studied is a double zone building. To reduce the computation time, the authors chose to apply the CFD simulation only in one zone and to pilot the system by imposing pressure as opening boundary conditions. The ESP-r simulation results provided boundary conditions to the CFD simulation.

Moreover, Srebric et al. [46] coupled a multizone tool called a ventilation simulating software CONTAM [47] with a CFD tool called PHOENICS-CFD [39] to evaluate the contaminant distribution in a building. First, they determined the airflow rates and the contaminant transport between zones. Then, they applied the CFD simulation only in the contaminant sources to deduce the airflow profile and the concentration distributions. These results are injected as fluxes in a new CONTAM simulation excluding the CFD domains. Finally, they evaluated the contaminant distribution. The authors showed that the coupled method is efficient in the zones very near the contaminant sources. However, in the other zones, the multizone approach remains the more appropriate method.

Finally, the CFD is particularly well adapted to describe flow fields in buildings. However, the large computation time makes difficult the generalization to all building applications. Indeed, in some cases, it is not necessary to give a very fine description and a way to overcome the difficulties enforced by the CFD is to model the building behaviour in a simpler manner by giving a less detailed description of the interested zone [48]. The first degree of CFD's simplification is the zonal technique. It is a way to obtain a more simple modelling while maintaining the complexity in a 2D map.

2.2. The zonal approach

2.2.1. Principle of the zonal approach

The zonal method is the first degree of simplification of the CFD technique. It has been introduced by Bouia and Dalicieux [49] and Wurtz [50] in the beginning of 1990s. This approach is a fast way to detail the indoor environment and to estimate a zone thermal comfort. Practically, it consists in dividing each building zone into several cells. One cell corresponds to a small part of a room. Therefore, the zonal method can be assumed to a two-dimensional approach. Fig. 2 represents a scheme in the case of zonal methods.

2.2.2. Advantages and application field of the zonal approach

The zonal formulation can treat a large volume space and the coupling between the system and its environment. The physical equations are solved for each cell of the zonal system. Consequently, it allows one to determine the local variables in a 2D-map. Thus, it is possible to evaluate the spatial distribution of different fields like temperature, pressure, concentration or air velocity remaining at a quite reasonable computation time. Wurtz et al. [51] showed that the zonal simulation is a suitable method for an accurate estimation of the temperature field in a room and of the indoor thermal comfort. Moreover, it allows also the visualization of building system airflows.

Several zonal modelling software in buildings are available. One of them frequently employed to describe and to visualize indoor airflows is the so-called SimSPARK software [52]. Equations are solved by the object-oriented environment called SPARK

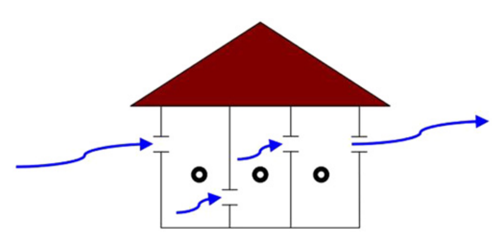

Fig. 3. Schematic representation of a problem solved with the multizone method (courtesy of Maxime Trocmé) [145].

[53]. Moreover, some researchers implemented their own zonal software as Haghighat with POMA [54].

2.2.3. Limitations of the zonal approach

As we mentioned above, the zonal approach is a minimization of the complexity of the CFD method. Thereby, it is obvious that some studies normally well implemented via the CFD are not anymore feasible via the zonal method [55–57]. Notably, some limitations reside in the following aspects:

- this technique requires previous knowledge on the flow profiles;
- it is not able to provide accurate results on the detailed description of the flow field;
- the study of the pollutant transport remains limited.

2.2.4. Applications reviews

Inard et al. [58] predicted the distribution of the air temperature inside a room with the zonal method. Especially, they proposed an original technique to model the mass air flow between two zones.

Musy et al. [59] studied the indoor thermal comfort in a room through the zonal software SPARK [53]. Particularly, the aim is to determine the vertical profile of temperature and the pollutant concentration repartition inside the room.

Tittelein et al. [60] focussed their works on the passive house and the methods to reduce the energy consumption of a building located in the region of Chambery, France. They compared the effects of a counter-flow ventilation and a single-flow ventilation on the energy efficiency.

Haghighat et al. [54] implemented a software using the zonal approach called POMA (Pressurized zOnal Model with Air-diffuser). This software allows one to predict the airflow pattern and the temperature distribution in a room which is naturally or mechanically ventilated.

Jiru and Haghighat [61] computed the airflow and the temperature distribution in a ventilated double skin facade, using the zonal method. Specifically, they compared the time evolution of the temperature in three positions inside the facade. Parametric studies have been accomplished in order to test the influence of the cavity height, the flow rate and the presence of venetian blinds on the inlet–outlet temperature difference.

Brun et al. [62] proposed both experimental and numerical studies to model heat transfers in a naturally ventilated roof cavity in timber-frame buildings in Grenoble, France. They used as numerical software the zonal software SPARK [53] to estimate the resulted heat gain considering the naturally ventilated cavity use.

Stephan et al. [20] were interested in inverse methods to improve the performance of the natural ventilation in a room. By coupling SimSPARK [52] with an optimization software called GenOpt [63], they deduced the optimal size of the openings needed to maximize the performance of the natural ventilation.

Abadie et al. [48] implemented an in-house zonal model in order to improve airflow modelling of forced convection in building zones. Especially, they were interested in developing an accurate model of the jet mass flow. This model was validated on a one-zone building.

The zonal technique particularly showed his efficiency in the description of flow profiles in building. However, such a detailed behaviour description is once again not always required and although it has been hugely enhanced with the zonal approach, the computation time can again be reduced by decreasing the complexity of the model. Thus, one more degree of simplification is proposed considering no more a multi-dimensional description of the building behaviour but a simple mono-dimensional visualization of the phenomenon occurring in the system.

2.3. The multizone or nodal approach

2.3.1. Principle of the nodal approach

This last approach, which is probably the simplest one, is called the multizone technique (also called nodal method). It considers the following assumption: each building zone is an homogeneous volume characterised by uniform state variables. Thus, one zone is approximated to a node that is described by a unique temperature, pressure, concentration, etc. Generally, a node represents a room, a wall or else the exterior of the building but it can be more specific like loads (internal occupancy or equipment gains, heating/cooling system). The thermal transfer equations are solved for each node of the system. In this term, the nodal method can be considered as a one-dimensional approach. Fig. 3 is a scheme of the nodal modelling.

TrnSys [64], EnergyPlus [43], IDA-ICE [65], ESP-r [45], Clim2000 [66,67], BSim [68,69] and BUILDOPT-VIE [70] are the most popular software using the nodal approach employed for building simulations.

In the literature, we can find two main methods used for the nodal approach: a first one consisting in solving transfer functions and a second one based on the finite difference method. Most software are designed from the first technique described by the transfer functions. The finite difference method is notably employed for nodal approaches using a description of the heat transfers from an electrical analogy. This technique has been introduced by Rumaniovski [71]. It is very useful since it simplifies drastically the physical problem through a linearization of the equations and thus, reduces the computation time. The principle of the electrical analogy is to associate a thermal resistance R and a thermal capacity C to a wall. The analogy gives the following equivalence with Ohm's law:

$$U_1 - U_2 = RI \Leftrightarrow \theta_1 - \theta_2 = \frac{e}{\lambda \cdot S} \Phi_L \tag{2}$$

The temperature θ is equivalent to voltage U, the heat flux Φ_L to current I and the thermal resistance $e/\lambda \cdot S$ to electrical resistance R. Several articles using this analogy have been published [72–79].

2.3.2. Advantages and application field of the nodal approach

The huge advantage of this technique resides in its ability to describe the behaviour of a multiple zone building on a large time scale with a small computation time. It is a particularly well-adapted tool for the estimation of the energy consumption and the time evolution of the space-averaged temperature into a room. Moreover, it can be used to predict the building air exchange rates and the airflow distribution between different rooms of a building. Some other applications as the ventilation efficiency or the pollutant transport for entire buildings can also be studied by this formulation [80,81]. Regarding the electrical analogy, additional advantages appear with the simplicity to implement the transfer equations and a more efficient computation time [78].

2.3.3. Limitations of the nodal approach

Due to the simplification enclosed in the multizone approach, it has obviously some limitations to investigate some specific cases [57,48], modelled more accurately and efficiently by the complete CFD method:

- The study of the thermal comfort and the air quality inside a zone is quite difficult.

- The impact of loads on their close environment is not addressed (for example, a radiator with a plume).
- Despite the fact that it is a well-adapted method to study a multiple zone building, it is quite difficult to apply the nodal form to a room with a large volume.
- Although it is a good way to visualize the distribution of pollutant between some building zones, it does not allow one to consider the local effects of a heat or pollutant source.

2.3.4. Applications reviews

Kalogirou [82] used a multizone software TrnSys (Transient Simulation Program) [64] to determine the energy consumption in a building in Nicosia, Cyprus. More specifically, the aim is to see how the energy demand behaves with a hybrid photovoltaic-thermal solar system (coupling of a normal PV panel and a heat exchanger) rather than a standard photovoltaic panel.

Ibanez et al. [83] used the TrnSys software to study the efficiency of the phase change materials (PCM) in Lleida, Spain. To perform that they considered a uniform indoor temperature in the room and determined its time evolution. By using the TrnSys software, they evaluated the influence of the PCM on different parts of the envelope of the room (wall, ceiling and floor).

Zhai et al. [84] studied the effects of the ventilation in summer on simulated data of indoor temperature with the multizone software EnergyPlus [43]. To achieve that they compared experimental and simulated measures of indoor temperature in three distinct building offices: a single-storey building with an automatically controlled air ventilation in Belgium, a three-storey building with a manually controlled air ventilation in Denmark and another three-storey building with an automatically controlled air ventilation in United Kingdom.

Cron et al. [75] used the electrical analogy to estimate the performance of hybrid ventilation. The system was composed of a fan-assisted natural ventilation incorporating a control demand strategy based on indoor air temperature and CO_2 concentration.

More recently, Bueno et al. [78] have developed an in-house resistance–capacitance model coupling the urban canopy with a building energy system. After a validation phase, they studied the effect of the urbanization on the energy consumption. Especially, they found a 5% increase of cooling systems in summer totally compensated by a 5% decrease of the heating during winter in residential buildings. Moreover, they were interested in the influence of the indoor environment on the outdoor air temperatures.

Goyal and Barooah [85] used the electrical analogy to implement a lumped thermal simulation model. It is able to predict the temperature and the humidity in multizone buildings from outside temperature and humidity, heat gains from occupants and solar radiation, supply air flow rates and supply air temperatures. Their objective was to decrease the order of this model by testing several reduction methods. Such scientific fields are really useful considering some applications such as HVAC control or monitoring.

Hazyuk et al. [79] developed an in-house multizone model from the electrical analogy. They proposed a description of the walls and the floor by two identical resistances and one capacity. The thermal mass is characterized by a single capacity and windows by single resistances. Having this kind of simplified model allows one to consider monitoring and control applications in more reasonable perspectives.

2.4. Discussion on the physical models

The previous paragraphs described several physical methods employed in the building modelling. We saw through the principles of each techniques and the previous examples that each physical method has its own application field. The most complete and detailed approach is the CFD. It allows one to describe very finely each mechanism occurring in the building system. Especially, it is particularly adapted for modelling the convective phenomenon taking place in a large zone volume. For instance, we saw through the examples of Zhai et al. [42], Wang and Wong [44] or Srebric et al. [46] the real necessity of using the CFD. Actually, in their study they treated very large volumes (office and auto racing complex) where the convective mechanisms are really complex. We mentioned above that the nodal approach assumes that the convection depends on the constant parameter h. So, it does not allow one to treat large zones with a high accuracy. Thus, in these specific cases, the use of the CFD was necessary. However, it is difficult to simulate all phenomenon by using the CFD because of the huge computation time. This is the reason why it is usually coupled with nodal software as EnergyPlus or TRNSYS. The nodal approach is really well adapted to treat global resolution as the determination of uniform field. Contrary to the CFD, phenomenon is described less finely. The aim is to simplify as far as possible the resolution system by linearizing the major part of the equations (when it is physically possible). Thus, the technical complexity is significantly reduced and both the computation time. For instance, Kalogirou [82] chose the nodal method because on the one hand, its studied system was constituted of several interconnected zones and on the other hand, he was interested in a specific macroscopic variable (energy consumption) and not in the distribution field. In the same way, Goyal and Barooah [85] and Hazyuk et al. [79] showed the necessity of the multizone method via the electrical analogy in control and monitoring perspectives. The zonal method is an intermediate technique between nodal and CFD approaches. It is

Table 1
Summary of the specificity of each physical technique.

Physical technique	Specificity of each technique	Application field	Advantages	Drawbacks
CFD method	One cell = a control volume (3-D); Local state variables	Contaminant distribution; Indoor air quality; HVAC systems	Detailed description of the fluid flows occurring inside the building; Large volume zones	Huge computation time; Complexity of the model implementation
Zonal method	One cell = a division of a room (2-D); Local state variables	Indoor thermal comfort; Artificial and natural ventilation	Spatial and time distribution of local state variables (temperature, concentration, pressure, airflow) in a large volume	Large computation time Requirement of a detailed description of the flow field and flow profiles
Nodal method	One cell = a room (1-D); Uniform state variables	Determination of the total energy consumption/the average of the indoor temperature/the cooling or heating load; Time evolution of the global energy consumption/ the space-averaged indoor temperature	Multiple zone buildings; Reasonable computation time; Easier implementation	Difficulty to study large volume systems Unable to study local effects as heat or pollutant source

less accurate than the CFD but retains more information compared to the nodal technique. As examples, Musy et al. [59], Tittelein et al. [60] or Haghighat [54] justified their choice of the zonal approach by the necessity to reduce the computation time compared with a CFD and the inability of the nodal method to provide detailed temperature and flow distribution and, in the same way, to predict the thermal comfort.

Moreover, all these techniques need some input parameters as meteorological data, geometrical data, thermo-physical variables or else occupancy and equipment scenario, etc. However, these parameters are always expressed under a certain part of uncertainties. Furthermore, in addition to these parameter uncertainties, there are also the uncertainties induced by the assumptions. Actually, several assumptions with consequences on the model performance have to be made in order to reduce the complexity of the thermal mechanisms occurring in buildings. Thus, all these uncertainties lead to a real difficulty to evaluate the accuracy degree of the models. Consequently, it seems very hard to gather all heat building transfers in a general overview without accumulating too much uncertainties [86].

We propose to gather in Table 1 the specificity of each method.

It comes out through the previous examples the need to reduce the computation time. Several solutions consisting in decreasing the system size exist and some of them have been described in the quoted articles [85,79]. Among the ideas not mentioned above, we can suggest also the building geometrical reduction by merging rooms or else merging walls. Such simplifications should speed up significantly the calculations and then, open to new application fields.

Generally, an important drawback of the physical formulation is the fact that it suggests a detailed description of the physical behaviour. Therefore, it implies expensive knowledge on the physical system, especially on the mechanisms occurring inside and outside the building geometry. Unfortunately, as we mentioned above, it is far from being always the case. In contrast, the statistical tools have the great faculty to product a model only from measures. Thus, we propose now to detail some statistical techniques frequently used in the building simulation and energy performances for prediction.

3. Statistical methods using machine learning

The particularity of statistical models compared with physical methods is the fact that they do not require any physical information. No heat transfer equations, no thermal or geometrical parameters are preliminary needed. Indeed, statistical models are based on the implementation of a function deduced only from samples of training data describing the behaviour of a specific system. Thus, these methods are well adapted when the physical features of the considered building are not known. Several statistical tools are able to build prediction model using learning methods. The great power of these techniques is the fact that they do not need to have much knowledge on the building geometry or the detailed physical phenomena to deduce an accurate prediction model. In contrast, they are totally based on measures and in such cases where it is difficult to collect data, it can become a real issue.

We propose in the following part of this paper to describe the statistical techniques mainly employed in the field of the building energy forecasting: the linear multiple regression, the artificial neural network, the genetic algorithm and the support vector machine. These techniques belong to the domain of the artificial intelligence.

Each sections are designed in the following manner: a first part describes succinctly the principle of the method, a second part shows the advantages and drawbacks of the method, a third part presents the application field and a last part gives some research applications using the method.

3.1. Multiple linear regression or conditional demand analysis (CDA)

The conditional demand analysis (CDA) is a linear multivariate regression technique applied to the building forecasting. The linear regression was introduced by Galton in 1886. In 1980, Parti and Parti were the first to propose a new method using the linear regression for the prediction of energy consumption in buildings: the conditional demand analysis [87]. The idea was to deduce the energy demand from the sum of several end-use consumption added to a noise term. In this way, they could infer the monthly and yearly residential end-use consumption from household invoices in San Diego.

3.1.1. Principle of the CDA

The principle of the linear multivariate regression is to predict Y as a linear combination of the input variables $(X_1, X_2, \ldots, X_p)$ plus an error term ϵ_i.

$$y_i = \alpha_0 + \alpha_1 \cdot x_{i1} + \alpha_2 \cdot x_{i2} + \cdots + \alpha_p \cdot x_{ip} + \epsilon_i, \quad i \in [1,n] \tag{3}$$

n is the number of sample data, p the number of variables and α_0 a bias. For example, if the predicted output is the internal temperature, there can be as inputs the external temperature, the humidity, the solar radiation and the lighting equipment.

3.1.2. Advantages and limitations of the CDA

The CDA technique can be used both for prediction or forecasting and for data mining. This method has a main advantage which is the simplicity of use by beginners since no parameter has to be tuned. Indeed, no specific expertise of the method is required to manage such type of prediction method.

However, the multiple linear regression presents a major limitation due to its inability to treat nonlinear problems. It leads to a lack of flexibility in forecasting but also a real difficulty to manage the multicollinearity inside the prediction results (that is the correlation between several variables). A possible solution to overcome these difficulties is to use a preliminary feature selection formulation.

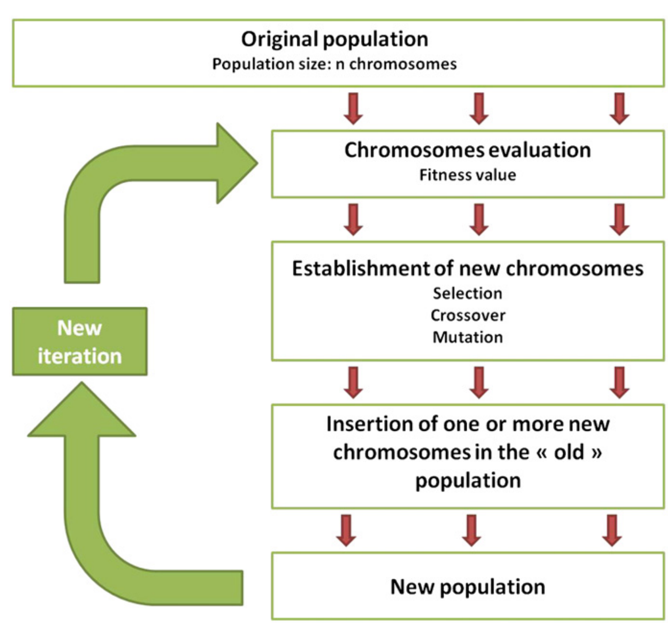

Fig. 4. Scheme of the general operation in the genetic algorithm.

3.1.3. Application field in CDA

In the building sector, the multivariable regression is often used for forecasting energy consumption or comparing the evolution of energy demand between two different periods. However, it is also employed for the prediction indoor air conditions, the control of HVAC equipment, reliability aspects and systems management [88,89]. The constraint is mostly present on data. Indeed, a large amount of data is required for a proper prediction and moreover the noncollinearity between variables is necessary [30].

3.1.4. Applications reviews

Lafrance and Perron [90] studied the evolution of the residential electricity demand at the regional level of Quebec in Canada. More specifically, they used the CDA as a signal processing tool and compared three years of data: 1979, 1984 and 1989.

Tiedermann [91] analysed the annual end-use consumption and the energy savings in the region of British Columbia in Canada. They studied also the energy consumption month by month and found two sudden increases: the first peak corresponds to November, December, January and February and is probably due to the use of the electric space heating and heating water. The second peak concerns the months of June, July and August and is related to the use of the air conditioning (central or portable).

Aydinalp-Koksal and Ugursal [30] used the CDA to model the residential end-use energy consumption in Canada at the national level. They kept their interest on several end-uses: appliances, lighting, space cooling, space heating and domestic hot water. Different energy sources have been studied: electricity, natural gas and oil. Each end-uses for each kind of energy were described by a linear regression.

More recently, Aranda et al. [92] has implemented a multiple linear regression model which allows one to predict the energy consumption in the banking sector in Spain and to suggest energy saving strategies to increase the energy efficiency. The authors chose a model able to combine the simplicity of the evaluation method and the accuracy of the result without needing a huge amount of input data.

Considering other applications, Givoni and Vecchia [93] proposed to use multivariate linear regression to describe the daily indoor average, minimal and maximal temperatures from outdoor measurements in occupied houses in Descalvado, Brazil. They found that all these temperatures can be predicted only from outdoor average, minimal and maximal temperatures. However, they showed that it is possible to improve the prediction of the indoor maximal temperature by adding the contribution of the solar radiation. Likewise, they improved the prediction of the minimal temperature by incorporating the dependence of the daily diurnal swings.

Few years later, Krüger and Givoni [94] showed that it is possible to estimate thanks to linear regression equations the indoor temperature behaviour in occupied low-cost houses in Curitiba, Brazil. More specifically, they linked linearly the average, minimal and maximal indoor temperatures to the average, minimal and maximal outdoor temperatures.

Nevertheless, due to the nonflexibility of the linear method, it is not always possible to use linear regression for all building applications. The following method is able to predict both linear and nonlinear problem. It is called genetic algorithm.

3.2. Genetic algorithm (GA)

The genetic algorithm (GA) is a stochastic optimization technique deduced from an analogy with the evolution theory of Darwin. This artificial intelligence method has been introduced in 1975 by Holland [95] but its use as an optimization tool for the building simulation started in the 1990's.

3.2.1. Principle of the GA

The principle of the genetic algorithm is based on the faculty of a given species to adapt itself to a natural environment and to survive extreme conditions. The genetic information is given by the gene sequences contained in the chromosome of an individual. In the GA process, all input variables are contained into one chromosome. This information can be coded in different way: binary, character string and tree. We will describe now the different step of the GA.

(1) Production of the original population.
(2) Evaluation of each chromosome based on the fitness value.
(3) Selection, crossover and mutation. The selection is responsible for selecting (at least) two chromosomes. After the selection step, the crossover phase can intervene, dealing with the exchange of a part of the information between the parents chromosomes. Then, the mutation operation can occur, consisting in the substitution of a part of a chromosome by another.
(4) Insertion of the new chromosomes in the population. At the end of the above processes (selection, crossover and mutation), some new chromosomes are added to the old population for creating the new one.
(5) Process reiteration. Once we have reach this step, the process restarts with the second step on the new population until the user specified generation number is completed.

Fig. 4 shows a scheme of the general operations in the genetic algorithm.

In the building simulation, GA is used to find a prediction model. The goal is to deduce a simple equation able to fit the problem. The form of the equation imposed by the user can have the following forms:

- linear $Y = w_1 \cdot X_1 + \cdots + w_n \cdot X_n$;
- quadratic $Y = w_1 \cdot X_1 + \cdots + w_n \cdot X_n + w_l \cdot X_1 \cdot X_2 + \cdots + w_m \cdot X_1 \cdot X_n + w_p \cdot X_2 \cdot X_3 + w_k \cdot X_n \cdot X_{n-1} + w_q \cdot X_1^2 + \cdots + w_r \cdot X_n^2 + w_s$;
- exponential $Y = w_1 + w_2 \cdot X_1^{w_3} + w_4 \cdot X_2^{w_5} + \cdots + w_l \cdot X_n^{w_{l+1}}$.

Y is the output (for example the energy demand), X_i are the input variables (for example the outdoor temperature, the humidity, the solar radiation and the exposure) and w_i are the weighting of each input variables. The GA is used to optimize the weighting w_i of each variables.

3.2.2. Advantages and limitations of the GA

An important advantage of genetic algorithm is the fact it deals with a powerful optimization method able to resolve every problems provided the convexity of the describing function [96]. Another essential advantage of the genetic algorithm is its ability to give several final solutions to a complex problem with a large number of input parameters. It allows the user to choose with his own judgement the most probable one. Obviously, this is also a drawback by the fact that the user can never be sure to have chosen the best solution, especially as the GA will not necessary generate the optimal solution. Another disadvantage of the GA is the large computation time. Some authors try to reduce this computation time by coupling the genetic algorithm with other statistical methods. Especially, Magnier and Haghigat [97] associated an artificial neural network to a genetic algorithm for estimating energy consumption and thermal comfort in a

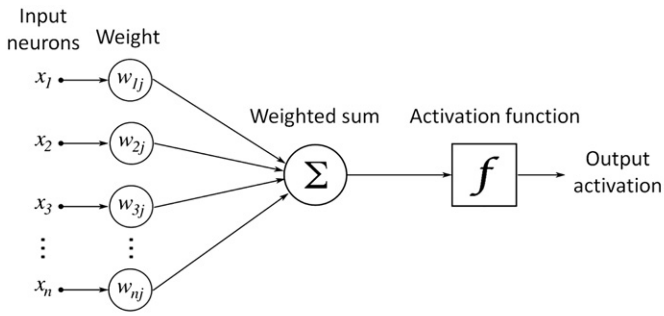

Fig. 5. Scheme of one neuron layer with the application of the activation function.

building. Another difficulty of the GA is the adjustment of the algorithm. Indeed, no rules are able to determine the number of individuals in the population, the number of generation or crossover and mutation probability. So, the only way to adjust the model is to test different combination. Another important limitation of the GAs is their capacity to generate local optimum leading to study the system locally instead of globally. Finally, the performance of the GA is really limited when the individuals present a similar evaluation value. In this case, the genetic algorithm can no longer evolve. Moreover, in this specific case, an important drawback is the fact that it is absolutely essential to postulate the form of the describing function.

3.2.3. Application field of the GA

In the building simulation, the genetic algorithm is mainly used for the determination of simple prediction models of the energy consumption and for the optimization of the equipment/load demand. The databases can be both simulated or real and can contain instantaneous samples on several time scale (hourly, monthly or yearly) or samples averaged in time and/or space. As the CDA, a large amount of data is required.

3.2.4. Applications reviews

Ooka and Komamura [98] were interested in the energy-efficiency in building during a day. With two genetic algorithms, they provided the optimized combination of equipment capacity and optimized operational planning for cooling system during a period of 24 h with an electric turbo refrigerator and a heat pump and water heating system with two distinct heat pumps for hot water. For the equipment capacity, the authors used an algorithm with a population size of 10 individuals (2 sub-populations with a size of 5 individuals), a number of generation of 30, a crossover probability of 1, a mutation probability of 0.01 and a migration probability of 0.5. For the operation planning, the GA presented a population size of 24 individuals (3 sub-populations with a size of 8 individuals), a number of generation of 750, a crossover probability of 1, a mutation probability of 0.01 and a migration probability of 0.5. This work was applied to an hospital of Tokyo in Japan on a period of 24 h.

Sadeghi et al. [99] used the GAs to implement optimized prediction models of the annual electricity consumption per inhabitant in residential sector in Iran. Three forms of simple equations are tested: linear, quadratic and exponential. Their variables are the annual gross domestic product, the annual real price of electricity and the annual real price of natural gas. The population size is 60 individuals, the number generation raises to 400, the probability crossover is equal to 0.5 and the probability mutation 0.02. The fitness was evaluated by the reverse of the sum squared error. Thus, the criterion was to maximize the fitness value. The selection process was the roulette-wheel method.

In the same manner, previous works of Ozturk et al. [100] studied the annual electricity consumption estimation in Turkey evaluated in the industrial sector and in the total electricity demand. The authors implemented two prediction models of the annual electricity consumption for both industrial and total turkish demand, allowing one to predict the annual electrical demand from 2002 to 2025.

Nevertheless, the genetic algorithms remain limited on a one hand by the choice of the parameters and the kind of function, and on the other hand by the large computation time and the uncertainty to obtain the optimal solution. The following technique overcomes all these difficulties. It is called the artificial neural network. Moreover, Datta et al. [101] compared the linear regression with the artificial neural network. They showed notably that this nonlinear technique performs quite better than the linear one. In the next part, we propose to introduce this technique and to give some examples from the literature.

3.3. Artificial neural network (ANN)

The artificial neural network (ANN) is a nonlinear statistical technique principally used for the prediction. This artificial intelligence method was inspired by the central nervous system with their neurons, dendrites, axons and synapses. It has been introduced in its mathematical form by McCulloch and Pitts in 1943. They published with Lettvin and Maturana the first works on the neural network in 1959 [102].

3.3.1. Principle of the ANN

The basic mono-layer ANN containing just two layers (input and output neurons) is described as the following steps:

(1) Choice of the inputs x_i considering the output(s). An initialization step associates each input to a weight w_i randomly chosen. The inputs are the neurons of the first layer.

(2) Application of the activation function f on the aggregation function. Most of the time, the aggregation function is a linear combination as

$$I = f\left(\sum_{i=0}^{n} w_i.x_i\right) \tag{4}$$

n is the number of input neurons and the product for $i=0$ is the bias. The activation function is responsible for converting the weighted input into an output activation. It returns a number between 0 and 1, allowing one to maintain the convergence (for example, sigmoid, Heaviside step or hyperbolic function). Fig. 5 shows a scheme of one neuron layer.

(3) Error calculation and application of the learning algorithm. The output is produced from the other steps. The global error corresponds to the sum of the training error calculated considering each data of the learning basis. To minimize the global error, a learning algorithm depending on a learning value is used to adjust the weight of each input neurons. The process is reiterated from step 2 to step 3 until reaching the error criterion.

3.3.2. Advantages and limitations of the ANN

An advantage of the ANN is that it does not need to detect the potential collinearity. Moreover, given their training faculty, another advantage of the ANN is its ability to deduce from data the relationship between different variables without any assumptions or any postulate of a model. Furthermore, it overcomes the discretization problem and is able to manage the data

unreliability. Finally, the ANN suggests a large variability of the predicted variable form (yes/no, binary 0 or 1, continuous value, etc.) and an efficient simulation time [103].

However, the ANNs are significantly limited by the fact that it implies to have a relevant database. Indeed, it is really important to train an ANN with an exhaustive learning basis with representative and complete samples (for example, samples in different seasons or in different moments of the day or during week-end/holidays, etc. and samples with each the same amount of information). Another disadvantage of the ANN is its large number of undetermined parameters (with no rules to determine them).

3.3.3. Application field of the ANN

In the building simulation, the artificial neural network are usually used for the prediction of the energy consumption or the forecasting of energy use as the cooling or heating demand without knowing the geometry or the thermal properties of the building. Different kinds of databases can be considered depending on the time scale as the hour, the month or the year and the nature of the data (real or simulated and instantaneous or time/space-averaged data). One main condition is absolutely essential for applying the artificial neural network technique: the completeness of the learning data. Kalogirou has published many works on the building applications using the ANN [103–105,82]. Particularly, in 2000, he presented a bibliographic review summing up the applications of the ANN in the field of energy-engineering systems [105].

3.3.4. Applications reviews

Kalogirou and Bojic [103] published a paper dealing with the prediction of the energy consumption of a passive solar holiday home in Cyprus during a day in summer and in winter. The inputs are the season, characteristics of the insulation, the masonry thickness, characteristics of the heat transfer coefficient and time of the day. The output is the energy consumption in kW h with a time-step of 10 min. The authors used a recurrent neural network containing four layers with 23 neurons on the hidden layers.

Aydinalp et al. studied the canadian annual electricity consumption in residential sector of appliances, lighting and cooling in a first paper (ALC) [28], and of space heating (SH) and domestic hot water (DHW) in a second paper [29]. In the first one, many inputs were used as appliances, weather, lighting, total heated area, socio-economic factors, etc. These information were propagated along a feed-forward network containing one input layer with 55 neurons, three hidden layers each of them including 9 neurons, and one output layer with one neuron representing the average of the annual electricity consumption due to the ALC.

Neto and Fiorelli [106] compared both an ANN model and a building software EnergyPlus for the forecasting of the energy demand in an administration building in Sao Paulo, Brazil. Two ANN model were tested: the first is a feed-forward neural network containing three layers: one input layer with 5 neurons (external temperature, humidity, two solar radiation parameters and day-type), one hidden layer with 21 neurons and one output layer with 1 neuron (daily total consumption). The second is a simpler ANN with only the external and internal temperature as inputs. The results for both simple ANN and complex ANN appeared to be very closed, indicating that the humidity and the solar radiation were certainly less significant than the external temperature for the forecasting of energy demand in this specific building study.

Recently, Kwok and Lee [107] studied the influence of the occupancy on the cooling load in Hong-Kong, China. They compared three different neural networks called probabilistic entropy-based neural network (PENN) to predict the total building cooling load: a first ANN containing 6 neurons on the external layer each of them characterizing a weather parameter, a second ANN with one more neuron (so 7 external neurons in total) for the hourly total occupancy area and a third ANN with another one more neuron (so 8 external neurons in total) corresponding to the occupancy rate (modification induced by the human presence). They found the best fitting between real data and the prediction for the last model (with 8 external neurons). It shows the huge influence of the occupancy on the building cooling load.

Moreover, Escriva-Escriva et al. [108] predicted the energy consumption based on building end-uses in University of Valencia, Spain. They used an ANN with multi-layer perceptron architecture consisting on three layers. The input layer contains four neurons (maximum temperature, minimum temperature, average temperature on just one day period and the average temperature on the day before), the hidden layer contains 3 neurons and the output layer consists in one neuron characterizing the energy consumption.

Recently, Leung et al. [109] used the artificial neural network to predict the cooling load in a university building in Hong-Kong. They took care of the occupancy by introducing a power demand. Thus, the input parameters are climatic data, hour and day type and pretreated air unit operation schedule. The output is the electrical power demand of the building cooling system. They used a feed-forward network with three layers. They found promising results especially when the cooling load is higher than the occupancy power demand.

However, the ANN is hugely limited by its lack of interpretability and the fact that it requires a large amount of learning data and mainly a relevant and completeness database (that is no missing data in the databases and the same amount of information for each variables). The following technique overcomes these difficulties given that it supports heterogeneous database and introduces a describing function. This method is called the support vector machine.

3.4. Support vector machine (SVM)

The support vector machine (SVM) has been introduced in 1995 by Vapnik and Cortes [110]. This artificial intelligence technique is usually used to solve classification and regression problems. Classification is a technique allowing one to divide a set of data in several categories, whose characteristics are given by the user. Regression method allows one to describe a set of data by a specific equation. The complexity of the regression equation is given by the user. We will focus our interest only on regression.

3.4.1. Principle of the SVM for regression

The principle of the SVM for regression is to find the optimal generalization of the model, in order to promote sparsity. Let us consider a given training data $[(x_1,y_1),\ldots,(x_n,y_n)]$, x_i being in the input space and y_i in the output space. In a nonlinear problem, the basic idea is to overcome the nonlinearity by transforming the nonlinear relation between x and y in a linear map. The way to do that is to send the nonlinear problem in a high-dimensional space called the feature space. As all regression techniques, the aim is to determine the function $f(x)$ that fits best the behaviour of the problem. The particularity of the SVM is the fact that it authorizes an error or an uncertainty ϵ around the regression function. The function f has the following form:

$$f(x) = \langle \omega, \Phi(x) \rangle + b \tag{5}$$

Φ represents a variable in the high-dimensional feature space and $\langle , \rangle$ a scalar product. ω and b are estimated by the following optimization problem called the primal objective function.

It corresponds to a minimization of the norm

$$\min_{\omega,b,\xi_i,\xi_i^*} \quad \frac{1}{2}\|\omega\|2+C\sum_{i=1}^{n}(\xi_i+\xi_i^*)$$
$$\text{subject to} \quad \begin{cases} y_i-\langle\omega,\Phi(x_i)\rangle-b\le\epsilon+\xi_i \\ \langle\omega,\Phi(x_i)\rangle+b-y_i\le\epsilon+\xi_i^* \\ \xi_i,\xi_i^*\ge 0 \end{cases} \tag{6}$$

C is a regularization parameter (a trade-off between the flatness of f and the maximal tolerated deviation larger than ϵ) imposed by users, ξ_i and ξ_i^* are two slack variables allowing a flexibility of the constraints. Moreover, by introducing a kernel function defined as a dot product in the feature space $k(x,x')=\langle\Phi(x),\Phi(x')\rangle$, it allows one to substitute a complex nonlinear map to a linear problem without having to evaluate $\Phi(x)$.

3.4.2. Advantages and limitations of the SVM

The main difficulty in the SVM is to select the best kernel function corresponding to a dot product in the feature space and the parameters of this kernel function. Some examples of kernel function mainly used in regression by SVM are given below:

- the linear kernel $k(x_i,x)=x_i\cdot x$;
- the polynomial kernel $k(x_i,x)=(x_i\cdot x+c)^d$;
- the radial basis function (RBF) kernel $k(x_i,x)=e^{\|x_i-x\|^2/2\sigma^2}$.

In addition to the kernel function parameters, two other constants have to be adjusted by users: the regularization constant C and the deviation ϵ.

The main advantage of the SVM is the fact that the optimization problem is based on the structural risk minimization principle (SRM). It deals with the minimization of an upper bound of the generalization error consisting of the sum of the training error. This principle is usually confronted to the empirical risk minimization (ERM) which only minimizes the training error. Another advantage is the fewer free parameters of optimization. Indeed, using the SVM technique required the adjustment of the regularization constant C and the margin ϵ. In contrast, the ANN method requires to know the topology of the inter-connections between neurons, the aggregation function, the number of hidden layers, the number of neurons on each hidden layers, the activation function, the learning algorithm (with the error calculation) and the learning value. In the same way, to implement a GA, we need to adjust the population size, the number of generation, the crossover probability and the mutation probability.

3.4.3. Application field of the SVM for regression

In building field, the SVM is mainly used for the forecasting of energy consumption or temperature. The system can be trained from different kinds of data with various time scales (year, month, hour) and various nature (instantaneous or space/time averaged). There is usually no restriction on the database except the fact that vector data are required. And a huge advantage is the fact that it supports a heterogeneous database that a database where all variables do not have the same amount of information or where we can find missing data.

3.4.4. Applications reviews

The use of support vector machine in the forecasting of energy consumption in buildings is quite recent. In 2005, Dong et al. [111] were the first to use SVM for the prediction of the building energy consumption. The aim is to predict the monthly energy consumption in four offices in Singapore. The input variables are the mean outdoor dry-bulb temperature, the relative humidity and the global solar radiation. The kernel function used is the radial basis function kernel.

Lai et al. [112] employed the SVM as a data mining tool for the prediction of the electrical consumption in residential sector in the region of Tohoku, Japan. Authors took as input parameters climate data as outdoor and indoor temperatures and humidities. They used the KXEN software [113] which consists in the implementation of the SVM method.

Li et al. [114,115] used the SVM in regression for the prediction of hourly cooling demand in Guangzhou, China. The aim is to predict the cooling demand hour by hour during summer in an office building. The input parameters are the outdoor dry-bulb temperature, the relative humidity and the global solar radiation. The SVM used as the kernel function a radial basis function.

Kavaklioglu [116] used the support vector regression method to predict the electricity consumption in Turkey until 2026. The kernel function is the radial basis function. The input variables are socio-economic parameters as population, Gross National Product, imports and exports.

Paniagua-Tineo et al. [117] employed support vector regression method to model and predict the daily air outdoor temperature in several European countries. The model depends on many prediction variables as the maximum and minimum temperature,

Table 2
Summary of the specificity of each statistical technique.

Statistical tool	Specificity of each technique	Application field	Advantages	Drawbacks
Conditional demand analysis: regression technique	Starting hypothesis: linear relation between variables and the output	Forecasting of the energy consumption; Evolution of the energy demand	Regression function describing the system	A large amount of training data/Non-collinearity between data
Genetic algorithm: optimization technique	Starting hypothesis: equation form imposed by the user; Final result is not necessary the best solution	Prediction of the energy consumption; Optimization of the equipment or load demand	Function describing the system; Powerful optimization algorithm	A large amount of training data; Difficulties to adjust algorithm parameters Large computation time
Artificial neural network: regression technique	No starting hypothesis but huge "black box" which prevents from physical interpretations	Prediction of the energy consumption and energy uses	A huge training faculty	A large amount of exhaustive and representative data; No physical interpretation
Support vector machine: regression technique	Starting hypothesis: kernel function imposed by the user	Forecasting of the energy consumption or temperature	A reasonable amount of training data with mainly vector data; Minimization problem based on the SRM	Determination of the kernel function Difficulty to adjust parameters C and ϵ

the precipitation, the relative humidity, the air pressure, the global radiation, the specific synoptic situation of the day and the so called monthly cycle. The kernel function is a Gaussian function.

Che et al. [118] proposed to develop an adaptive fuzzy rule based prediction system combining the SVM in regression and a fuzzy inference method with the aim to forecast the electrical load in New South Wales. The authors used the radial basis function as kernel function.

Chen et al. [119] estimated the monthly mean daily solar radiation in Chongqing, China via the support vector machine method. More particularly, the aim is to improve the state of data collected in the station. The authors chose to test three different kernel function: linear, polynomial and radial basis function. Also, they proposed to experiment seven combinations of input variables only based on the maximum temperature and the minimum temperature. Finally, they implemented 21 different SVM system.

3.5. Discussion on the statistical tools

Contrary to the physical techniques which are each associated with a specific application, we realize that no statistical tool has a better use for a problem than another. However, it is possible to classify them by complexity. Indeed, the linear multiple regression is probably the easier statistical method. It is able to give good prediction and does not need a real expertise to be implemented. But it is hugely limited by the fact that it assumes a linear description of phenomenon. The genetic algorithm is a bit less limited because it is able to treat both linear and nonlinear problems. But it suggests that the function describing the system behaviour is well-known. However, it is rarely the case. Moreover, another huge limitation of the genetic algorithm is the choice of input parameters. The artificial neural network overcomes this problem given that it does not need to give specific description. Nevertheless, it runs as a black-box system which makes the interpretability very difficult. Moreover, an important drawback of the ANN is the fact that it requires a large amount and a completeness of learning data. In contrast, the Support Vector Machine has the huge advantage to do not need completeness data. And due to the known kernel function, the problem remains interpretable. However, contrary to the artificial neural network, it requires to assume the form of the kernel function. Finally, we see that each of these statistical techniques has his own advantages and drawbacks and the choice of the method depends mainly on the user and on what he expects at the end of the study. Therefore, the technique can be chosen according to the targeted outcome. We propose to sum up the specificity of each statistical technique in Table 2.

4. Hybrid models

The previous parts of this paper showed the capacity of both detailed physical and statistical methods in the building simulation. But they showed also the limitations of each techniques. Especially, the white box methods assume that all building characteristics, both thermal and geometric one, are well-known. This is usually the case for building design but it is more difficult to collect so many information on existing buildings. However, to establish monitoring strategies, they are absolutely required. Moreover, these approaches suggest that we are able to describe all physical mechanisms with a high accuracy. Nevertheless, although most of the thermal phenomenon are well-known, some of them are based on assumptions and remain difficult to model accurately as the natural ventilation often described by empirical equations. The black box methods are mainly limited by the fact that they absolutely required data and mostly in large amount. Moreover, it is usually difficult to interpret results obtained by statistical approaches in physical term. Otherwise, data mining techniques are specific to a building. Thus, the treatment of another building leads to a new modelling. In contrast, due to the general heat transfer equations, white box methods are usually applied generally.

It is possible to overcome the limitations of each technique by coupling them. Indeed, the advantages of a method remove the drawbacks of the other one. For example, by retaining a part of physical meaning, one keeps always the interpretability of the problem. Moreover, building characteristics can be determined by optimization techniques such as genetic algorithms. Thus, all physical and geometrical input parameters are not any more required. These hybrid methods combining physics and statistics are called “grey box” methods.

4.1. Principle of the hybrid approach

Generally speaking, the principle of the hybrid methods is based on the coupling of statistical methods and physical models. In this way, several strategies are available.

A first strategy consists in using machine learning as physical parameters estimator. We will see in the following examples that most of the time, scientists couple a nodal model with genetic algorithms.

A second strategy is to use statistics in order to implement a learning model describing the building behaviour. This learning model is designed from a learning basis built from a physical approach. In the following, we will present some examples employing this technique.

A third strategy consists in using statistical method in fields where physical models are not effective and accurate enough. For example, end-uses are known to be really difficult to take into account in physical models. In contrast, statistical techniques allow one to well-consider these end-uses. So, a solution would be to associate both physical and statistical methods in order to implement the complete system. Another application would be to determine the heat behaviour in a multiple zones building where the thermal properties of some rooms would be unknown. Thus, some zones could be physically studied while others would need to be described statistically via measurements collected in these zones. This strategy is currently not referenced in the building simulation literature. However, it has already been proven in other fields as the prediction of the battery behaviour [120].

4.2. Advantages and limitations of the hybrid methods

The main advantage of the hybrid method is that it allows one to consider only a limited number of data. Furthermore, the input parameters do not need to be fixed at the initial time of the simulation. Only bounds on physical parameters are required. Thus, a rough description of the building geometry and thermal parameters is sufficient. Also, the hybrid methods allow one to retain a physical interpretation.

However, some drawbacks own to each technique remain in the hybrid method as the free parameters for statistical tool or the computation time needing for both physical or statistical codes.

A last drawback that is also an advantage is the fact that the grey box method couples two distinct scientific domains. Although it brings some difficulties for users to understand, it should be of a great scientific interest.

4.3. Application field of the hybrid method

This approach has been introduced at the beginning of the 1990s for a specific application which was the automatic control system. For example, Teeter and Chow [121] combined an artificial neural network with a single-zone thermal model to improve the efficiency of the HVAC control by performing the HVAC parameters identification. Other more recent examples are the works of Paris et al. [12,122] who combined the fuzzy logic, a PID controller and a dynamic model describing the thermal behaviour of the building for implementing several heating control schemes. Furthermore, Nassif et al. [123] applied an optimization process to HVAC system for monitoring issues. It consisted in identifying the zone air temperatures, the supply air temperature, the supply duct static pressure, the zone supply air temperature or reheat required, the minimum outdoor ventilation flow rate, and the chilled water supply temperature. In the same applications, we can also point out the work of Caldas and Norford [96] on the control of HVAC systems.

As we mentioned above, another application of the hybrid model is the parameters identification. In this approach, the aim is to compute the set of input values corresponding to a given set of outputs. For instance, the objective may be to calculate the optimal thermal properties of the walls (conductivity, capacity, etc) given a target consumption/comfort level. The technique is to combine physical models – used to simulate the thermal behaviour of the building – and statistical technique to retrieve the set of optimal inputs corresponding to the desired outputs.

Concerning the amount of data required, it is quite reasonable by the fact that it includes a part of physical interpretation inside the program.

In the literature related to this topic, some papers focus on the coupling between nodal techniques for the thermal and geometrical representation and genetic algorithms for the parameters identification. Others deal with the coupling between regression techniques and thermal models. We propose to give some applications using these two kinds of hybrid methods.

4.4. Applications reviews

As we mentioned above, a frequently used hybrid technique consists in coupling thermal building model with genetic algorithms for parameters identification. More precisely, a given number of set of parameters is produced by genetic algorithm and each of them is tested on the thermal model. The fitness value is then evaluated as the error of the model output. The set of parameters giving the smallest fitness value is, then, the best solution.

Lauret et al. [124] implemented a model resolving the state equations in a building with a very simple geometry in the Island of the Reunion in order to follow the evolution of the indoor dry air temperature. To do that they combined a physical resolution by the finite difference method via the multizone software CODYRUN [125] with a genetic algorithm. The study is based on the experimental data. The authors have shown in previous studies that the physical model alone was not allowed us to return a good agreement with the real data [72,126]. In this study, they used the genetic algorithm to isolate the defective node measurement by forcing the value of some temperatures in specific place of the building. The aim is to optimize the value of the indoor dry air temperature.

Znouda et al. [127] studied energy consumption in a Mediterranean building in Tunisia. More specifically, they found the solutions to improve both the energy efficiency and the economic point of view by optimizing architectural parameters. To perform that they coupled a simplified tool for building thermal evaluation specific to the Mediterranean countries called CHEOPS [128] to a genetic algorithm for the architectural parameters identification. They studied the energetic and economic problem independently. The authors studied a solution adapted both in summer and in winter. They showed that it is difficult to solve this kind of multi-objective issue composed of two independent problems (energetic and economic) because the optimal solutions are different considering either saving energy or saving money.

Wang and Xu [129,130] studied the building thermal transfer in summer in Hong-Kong. The building consists of three different buildings one of them containing offices, another a shopping center and the last one a restaurant. The study was based on data collected during a survey conducted in order to deduce the profile of occupancy and use of the lighting and equipment. They used the electrical analogy to predict the heating/cooling load by substituting the building envelope (the roof and the external wall) by two different 3R2C systems and by introducing an internal mass by a 2R2C system. The internal mass corresponds to all others heat storage materials as furnitures, carpet, partitions, equipment, etc. Combining the equations resolution with the genetic algorithm for the parameter identification, the authors optimized the values of the resistances and capacitances of the internal mass.

Tuhus-Dubrow and Krarti [131] implemented an hybrid model by combining the nodal software DOE-2 [132] with genetic algorithms in order to determine the most efficient building shape considering different parameter sets and output's criterion. Among rectangle, U-shape, H-shape, T-shape, L-shape, cross-shape and trapezoidal buildings, the rectangle and trapezoidal one were the most efficient shapes both in term of energy consumption and life-cycle cost. Nevertheless, variations between all studied shapes were quite small allowing one to give a large flexibility to architects.

Siddarth et al. [133] have coupled genetic algorithm and DOE-2 [132] in order to establish a database allowing them to implement regression functions describing the annual energy consumption. Indeed, they used genetic algorithm to generate several set of parameters. Each set of parameters has been tested in DOE-2 [132] which returned the annual energy consumption. Part of these set of parameters are then selected under an annual energy consumption criterion and injected inside a database, which will be used for the implementation of a regression

Table 3
Comparison between white, black and grey box techniques.

Methods	Building geometry	Training data	Physical interpretation
Physical or "white box" method	A detailed description of the building geometry is required	No training data are required	Results can be interpreted in physical terms
Statistical or "black box" method	A detailed description of the geometry is not required	A large amount of training data collected over an exhaustive period of time is required	There are several difficulties to interpret results in physical terms
Hybrid or "grey box" method	A rough description of the building geometry is enough	A small amount of training data collected over a short period of time is required	Results can be interpreted in physical terms

function. Under this annual energy consumption model, they are, thus, allowed one to suggest energy saving strategies.

Sahu et al. [134] proposed a strategy consisting to couple electrical analogy model with a genetic algorithm for improving design building parameters to reduce plant load. More specifically, they identified the orientation, the shape, the roof and walls materials and window properties. They validated their results by comparing the model response with the commercial software TrnSys [64].

Yang et al. [135] tested several evolutionary algorithms to identify building parameters for energy savings. More specifically, they coupled the software HAMbase [136] with these algorithms to optimize external and internal wall properties as the thermal resistances and capacities, and also long-wave and short-wave radiation coefficients as the emissivity or the absorptivity. Their objective was to minimize the fitness value defined as the mean absolute error.

A second technique proposed in several articles was introduced in the 1990s by Lam et al. [137,88]. They suggested a new strategy consisting in generating a database from a thermal dynamic simulation software. Those data are then used as input parameters in a regression tool. Several techniques can be used as regression techniques as multivariate regression, artificial neural network and support vector machine. The advantage is that it is possible to predict outputs from the regression equations without needing to resort to the simulation building software. Thus, this technique allows one to reduce significantly the computation time.

In this specific case, Lam et al. [88] used the nodal software DOE-2 [132] as database generator and implemented a prediction model from multivariate linear and nonlinear regression equations able to find the annual energy consumption function of 12 selected variables in air-conditioned office building in Hong-Kong.

Likewise, Freire et al. [89] proposed a strategy consisting in generating a database from their in-house model called Power-Domus [138]. Those data are used as input parameters in a regression tool. Particularly, they were interested in predicting the indoor temperature and the relative indoor humidity from the outdoor temperature, the relative outdoor humidity, the total solar radiation, the heating load and the HVAC power.

In the same way, Xu et al. [139] established a model coupling the nodal software EnergyPlus [43] with an artificial neural network for predicting the energy consumption. More specifically, they generated a database from the thermal model that they put as input parameters of the ANN. After training it, the ANN was able to deduce the prediction of the energy consumption.

More recently, Lee et al. [140] proposed to couple a regression analysis with a thermal simulation model to describe the influence of the size, thermal properties and orientation of windows in buildings considering 5 different climate zones in Asia. Their main goal was to deduced optimized parameter windows able to reduce the cooling or/and heating load.

4.5. Discussion on the hybrid methods

Through the previous examples, it appears that the hybrid method is mainly used for parameters estimation. Contrary to statistical or physical approaches that we described above, the aim is not any more just to predict the thermal behaviour of a specific building but mainly to return different strategies able to improve energy efficiency. That is why, it is particularly well adapted for the monitoring issues. Indeed, the hybrid technique selects the advantages of both physical and statistical methods and uses them to implement efficient models for monitoring and control applications. Thus, it allows one to keep a part of physical interpretation while not requiring a really accurate description of all phenomena occurring in the heat building transfer.

The hybrid method is also a remarkable scientific challenge by the fact that it implies several scientific domains as physics and statistics. Indeed, it promotes the collaboration between these two disciplines. Indeed, we saw two specific techniques combining machine learning and thermal modelling. However, in the future, thanks to both statistician and physician experts and especially their ability to work together, the grey method could be extended to other new combinations models. Actually, this scientific field being relatively recent, lots of improvements in the hybrid method must still be accomplished.

Considering these promising perspectives, our team took recently part in this scientific field by developing our own hybrid model. We propose to couple a simplified in-house thermal model based on the electrical analogy with a multivariate regression to create several metamodels from an initial database designed from the thermal model. This method has already be tested for other specific applications [141–143]. Preliminary results are really promising concerning the feasibility of the method in building applications.

Most of these works present results validated on specific cases. An interesting outcome would be to find some generic regression equations. The main issue is that it probably requires a large amount of parameters. Larger the parameters quantity, larger the database and larger the resort to computation software and by this way the computation time. Thus, it is a really interesting issue since it is a great scientific challenge that could have a remarkable impact.

We propose to sum up in Table 3 the properties of the hybrid techniques compared with those of physical and statistical tools.

5. Conclusion

In this paper, we have proposed a review of the main techniques and tools enabling building energy performances prediction. These techniques have been introduced along three categories, each of them associated to specific scientific paradigms and fields: First of all, approaches relying on physical models (“white box” methods) have been introduced. These may be divided into three sub-categories, which mainly correspond to a gradual rise of the level of details of building models: the multizone technique which considers the space as a homogeneous volume where all states variables are uniform, the zonal method which divides each room in several cells and the CFD method which describes each zones in several control volumes. Then, we have focused on methods based on machine learning (or ‘‘black box” methods), which rely on statistical treatments of building energy and comfort data. Four methods have been reviewed: conditional demand analysis, artificial neural networks, genetic algorithms and support vector machine. The last category of methods considered is the one of hybrid approaches which rely on both physical models in order to simulate building thermal behaviour and machine learning techniques in order to optimize input parameters. Finally, a critical synthesis has been performed in order to highlight for each method the most appropriate applications. The first kind of methods – those relying on physical models – are mostly applicable to contexts in which building design data are available, and especially in the scope of the design of a new building. Actually, those methods rely on quite detailed descriptions of buildings, notably entailing geometry, material properties, and energy systems features. While this information can be considered to be easily extractable from design data in the case of a new building, this is less than obvious for existing buildings (e.g. in the scope of a refurbishment). This is true for the most basic of these methods – the nodal one – but all the more true when we consider more advanced ones (zonal, CFD). When it

comes to comparing those methods between them, the conclusion is quite straightforward: obviously, it is better to use more detailed models (CFD) for the sake of reliability and precision of simulation result, but models are more tedious to build and computation times are higher. Zonal methods can be considered as good trade-offs, but still, most simulation tools used today in "real-life" projects are based on nodal approaches. Nevertheless, a possible trend is a gradual shift to CFD methods with computers becoming more powerful. The second category of methods, which are based on machine learning techniques, are extremely useful in opposite situations, i.e. those in which one owns real energy and comfort data from the building but has little or no information about the design. But the reliability of these techniques is highly dependent on the quality and amount of available data, as were the physical approaches dependent on the complexity of the underlying model. It is however quite difficult to perform a qualitative and comparative assessment of the various techniques devised in this field, since – again – their performances will depend on the training data used as input. Compared to physical approaches, machine learning ones require less information about the building and may appear as easier to deploy. However, physical approaches are more handy in scopes where interpretation of physical phenomena is desired. At last, hybrid approaches appear as a very promising field for the near future [144]. They can be considered as a nice trade-off between physical and machine-learning based methods, and relax their drawbacks by combining them. Hybrid methods may be appreciated in situations were a building physical model is available, but is incomplete or does not offer enough details, and therefore has to be adapted and/or completed. When dealing with existing buildings, where it is usually difficult to rebuild detailed physical model, such approaches could be of great help.

References

[1] EPBD. On the energy performance of buildings. Official Journal of the European Union, Directive 2010/31/EU of the European Parliament and of the Council; 2010.

[2] Poel Bart, van Cruchten Gerelle, Balaras Constantinos A. Energy performance assessment of existing dwellings. Energy and Buildings 2007;39:393–407.

[3] Balaras Constantinos A, Gaglia Athina G, Georgopoulou Elena, Mirasgedis Sevastianos, Sarfidis Yiannis, Lalas Dimitris P. European residential buildings and empirical assessment of the hellenic building stock, energy consumption, emissions and potential energy savings. Building and Environment 2007;42:1298–314.

[4] Gaglia Athina G, Balaras Constantinos A, Mirasgedis Sevastianos, Georgopoulou Elena, Sarafidis Yiannis, Lalas Dimitris P. Empirical assessment of the hellenic non-residential building stock energy, consumption emissions and potential energy savings. Energy Conversion and Management 2007;48:1160–1175.

[5] Proposal-EPBD. Low energy building in Europe: current state of play, definitions and best practice; 2009.

[6] Jean-Loup Bertez, The PASSIVE stake. Strategic overview on a global, structured and sustainable way for "efficient building". Zenergie; 2009.

[7] Barlow Stuart, Fiala Dusan. Occupant comfort in UK offices—how adaptive comfort theories might influence future low energy office refurbishment strategies. Energy and Buildings 2007;39:837–46.

[8] Ed Carroll, Eric Hatton, Mark Brown. Residential energy use behavior change pilot. CMFS Project Code B21383. Franklin Energy; 2009.

[9] Badescu Viorel, Sicre Benoit. Renewable energy for passive house heating. Part I. Building description. Energy and Buildings 2003;35:1077–84.

[10] Badescu Viorel, Sicre Benoit. Renewable energy for passive house heating II. Model. Energy and Buildings 2003;35:1085–96.

[11] Bohm B, Danig PO. Monitoring the energy consumption in a district heated apartment building in Copenhagen, with specific interest in the thermodynamic performance. Energy and Buildings 2004;36:229–36.

[12] Paris Benjamin, Eynard Julien, Grieu Stphane, Talbert Thierry, Polit Monique. Heating control schemes for energy management in buildings. Energy and Buildings 2010;42:1908–17.

[13] Petersen Steffen, Svendsen Svend. Method for simulating predictive control of building systems operation in the early stages of building design. Applied Energy 2011;88:4597–606.

[14] Aswani Anil, Master Neal, Taneja Jay, Culler David, Tomlin Claire. Reducing transient and steady state electricity consumption in HVAC using learning-based model-predictive control. Proceedings of the IEEE 2012;100(1).

[15] Mossolly M, Ghali K, Ghaddar N. Optimal control strategy for a multi-zone air conditioning system using a genetic algorithm. Energy 2009;34:58–66.

[16] Zhao Hai-xiang, Magoulès Frédéric. A review on the prediction of building energy consumption. Renewable and Sustainable Energy Reviews 2012;16:3586–92.

[17] Liao Z, Dexter AL. A simplified physical model for estimating the average air temperature in multi-zone heating systems. Building and Environment 2004;39:1013–22.

[18] Badescu Viorel. Simple and accurate model for the ground heat exchanger of a passive house. Renewable Energy 2007;32:845–55.

[19] Li Yuguo, Delsante Angelo, Symons Jeff. Prediction of natural ventilation in buildings with large openings. Building and Environment 2000;35:191–206.

[20] Stephan Louis, Bastide Alain, Wurtz Etienne. Optimizing opening dimensions for naturally ventilated buildings. Applied Energy 2011;88:2791–801.

[21] Wang Shengwei. Dynamic simulation evaluation of building VAV air-conditioning system and of EMCS on-line control strategies. Building Simulation 1999;34:681–705.

[22] Ihm Pyonchan, Nemri Abderrezek, Krarti Moncef. Estimation of lighting energy savings from daylighting. Building and Environment 2009;44:509–514.

[23] Clarke JA, Johnstone C, Kelly N, Strachan PA. The simulation od photovoltaic-integrated building facades. Building Simulation 1997:189–95.

[24] Chow TT, Hand JW, Strachan PA. Building-integrated photovoltaic and thermal applications in a subtropical hotel building. Applied Thermal Engineering 2003;23:2035–49.

[25] Ordenes M, Lamberts R, Guths S. Estimation of thermophysical properties using natural signal analysis with heat and moisture transfer model. Energy and Buildings 2009;41:1360–7.

[26] Qin Menghao, Walton George, Belarbi Rafik, Allard Francis. Simulation of whole building coupled hygrothermal airflow transfer in different climates. Energy Conversion and Management 2011;52:1470–8.

[27] Entrop AG, Brouwers HJH, Reinders AHME. Evaluation of energy performance indicators and financial aspects of energy saving techniques in residential real estate. Energy and Buildings 2010;42:618–29.

[28] Aydinalp Merih, Ugursal V Ismet, Fung Alan S. Modeling of the appliance, lighting, and space-cooling energy consumptions in the residential sector using neural networks. Applied Energy 2002;71:87–110.

[29] Aydinalp Merih, Ugursal V Ismet, Fung Alan S. Modeling of the space and domestic hot-water heating energy-consumption in the residential sector using neural networks. Applied Energy 2004;79:159–78.

[30] Aydinalp-Koksal Merih, Ugursal V Ismet. Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector. Applied Energy 2008;85:271–96.

[31] Gugliermetti F, Passerini G, Bisegna F. Climate models for the assessment of office buildings energy performance. Building and Environment 2004;39:39–50.

[32] Wall M. Distribution of solar radiation in glazed spaces and adjacent buildings. A comparison of simulation programs. Energy and Buildings 1997;26:129–35.

[33] McDowell Timothy P, Thornton Jeff W, Emmerich Steven, Walton George. Integration of airflow and energy simulation using CONTAM and TRNSYS. ASHRAE Transactions 2003;109 KC-03-10-2.

[34] Crawley Drury B, Hand Jon W, Kummert Michael, Griffith Brent T. Contrasting the capabilities of building energy performance simulation programs. Building and Environment 2008;43:661–73.

[35] Woloszyn Monika, Rode Carsten. Tools for performance simulation of heat, air and moisture conditions of whole buildings. Building Simulation 2008;1:5–24.

[36] Adrien Brun, Clara Spitz, Etienne Wurtz, Laurent Mora. Behavioural comparison of some predictive tools used in a low-energy building. In: Building simulation 11th international IBPSA conference; 2009. p. 1185–90.

[37] 〈http://www.ansys.com/Products/Simulation+Technology/Fluid+Dynamics/ANSYS+Fluent〉; 2012.

[38] 〈http://www.comsol.com/〉; 1998.

[39] CHAM. Phoenics user manual for program version 3.6. CHAM Ltd.; 2005.

[40] Tan Gang, Glicksman Leon R. Application of integrating multi-zone model with CFD simulation to natural ventilation prediction. Energy and Buildings 2005;37:1049–57.

[41] Qin Da Yan Rong, Zhou Yi Jiang Xin. Research on a dynamic simulation method of atrium thermal environment based on neural network. Building and Environment 2012;50:214–20.

[42] Zhai Zhiqiang, Chen Qingyan, Haves Philip, Klems Joseph H. On approaches to couple energy simulation and computational fluid dynamics programs. Building and Environment 2002;37:857–64.

[43] EnergyPlus. 〈www.eere.energy.gov/buildings/energyplus〉, U.S. Department of Energy (DOE).

[44] Wang Liping, Wong Nyuk Hien. Coupled simulations for naturally ventilated residential buildings. Automation in Construction 2008;17:386–98.

[45] Clarke JA, McLean D. ESP—a building and plant energy simulation system. Energy Simulation Research Unit University of Strathclyde, Strathclyde; 1988.
[46] Srebric Jelena, Yuan J, Novoselac Atila. On-site experimental validation of a coupled multizone and CFD model for building contaminant transport simulations. ASHRAE Collections NY-08-033; 2008.
[47] Dols WS, Walton GN. CONTAMW2.0 user manual. National Institute of Standards and Technology (NIST); 2002.
[48] Abadie MO, de Camargo MM, Mendona KC, Blondeau P. Improving the prediction of zonal modeling for forced convection airflows in rooms. Building and Environment 2012;48:173–82.
[49] Bouia H, Dalicieux P. Simplified modelling of air movements inside dwelling room. In: Proceedings of building simulation 91 conference, Nice, France, IBPSA (The International Building Performance Simulation Association); 1991. p. 106–10.
[50] Wurtz Etienne. Three-dimensional modeling of thermal and airflow transfers in building using an object-oriented simulation environment [in French]. PhD thesis, Ecole Nationale des Ponts et Chaussees; 1995.
[51] Wurtz Etienne, Mora Laurent, Inard Christian. An equation-based simulation environment to investigate fast building simulation. Building and Environment 2006;41:1571–83.
[52] Mora Laurent. Prédiction des performances thermo-aérauliques des bâtiments par association de modèles de différents niveaux de nesse au sein dun environnement orienté objet. PhD thesis, Université de La Rochelle; 2003.
[53] LBNL. SPARK 2.0 reference manual. Lawrence Berkeley National Laboratory and Ayres Sowell Associates Inc.; 2003.
[54] Haghighat Fariborz, Li Yin, Megri Ahmed C. Development and validation of a zonal model POMA. Building and Environment 2001;36:1039–47.
[55] Inard C. Contribution à létude du couplage thermique entre un émetteur de chauffage et un local. Etudes expérimentales en chambres climatiques. PhD thesis, National Institute of Applied Sciences (INSA), Lyon, France; 1988.
[56] Wurtz Etienne, Nataf Jean-Michel, Winkelmann Frederick. Two- and three-dimensional natural and mixed convection simulation using modular zonal models in buildings. International Journal of Heat and Mass Transfer 1999;42:923–40.
[57] Mora L, Gadgil AJ, Wurtz E. Comparing zonal and CFD model predictions of isothermal indoor airflows to experimental data. Indoor Air 2003;13:77–85.
[58] Inard Christian, Bouia Hassan, Dalicieux Pascal. Prediction of air temperature distribution in buildings with a zonal model. Energy and Buildings 1996;24:125–32.
[59] Musy Marjorie, Winkelmann Frederick, Wurtz Etienne, Sergent Anne. Automatically generated zonal models for building air ow simulation: principles and applications. Building and Environment 2002;37:873–81.
[60] Tittelein P, Wurtz E, Achard G. SimSpark platform evolution for low-energy building simulation. International Scientific Journal for Alternative Energy and Ecology 2008;62.
[61] Jiru Edae Teshome, Haghighat Fariborz. Modeling ventilated double skin facade—a zonal approach. Energy and Buildings 2008;40:1567–76.
[62] Brun Adrien, Wurtz Etienne, Quenard Daniel. Experimental and numerical comparison of heat transfer in a naturally ventilated roof cavity. In: Building simulation fourth national conference of IBPSA-USA; 2010. p. 160–9.
[63] Michael Wetter. GenOpt—generic optimisation program. Berkeley: Lawrence Berkeley National Laboratory; 2008.
[64] Klein KA. A transient simulation and program. Madison, WI: Solar Energy Laboratory; 1996.
[65] 〈www.equa.se〉; 2008.
[66] Bonneau D, Rongere FX, Covalet D, Gauthier B. Clim2000: modular software for energy simulation in buildings. In: Proceedings of IBPSA 93 Adelaide, Australia; 1993.
[67] Woloszyn M, Rusaouen G, Covalet D. Whole building simulation tools: Clim2000. Publication A41-T1-F-04-3. Presentation for IEA Annex 41 meeting Zurich, Switzerland; 2004.
[68] Rode C, Grau K. Whole building hygrothermal simulation model. ASHRAE Transactions 2003;109:572–82.
[69] Rode C, Grau K. Integrated calculation of hygrothermal conditions of buildings. Presentation for IEA Annex 41 meeting, Zurich, Switzerland, Publication A41-T1-DK-04-1; 2004.
[70] BuildOpt-VIE. 〈www.bph.tuwien.ac.at〉, University of Technology of Vienna; 2007.
[71] Rumaniovski P, Brau J, Roux J-J. An adapted model for simulation of the interaction between a wall and the building heating system. In: Poceedings of the thermal performance of the exterior envelopes of buildings IV conference Orlando, USA; 1989. p. 224–33.
[72] Mara Thierry Alex, Garde Franois, Boyer Harry, Mamode Malik. Empirical validation of the thermal model of a passive solar cell test. Energy and Buildings 2001;33:589–99.
[73] Fraisse Gilles, Viardot Christelle, Lafabrie Olivier, Achard Gilbert. Development of a simplified and accurate building model based on electrical analogy. Energy and Buildings 2002;34:1017–31.
[74] Gouda MM, Danahera S, Underwood CP. Building thermal model reduction using nonlinear constrained optimization. Building and Environment 2002;37:1255–65.
[75] Cron Florence, Inard Christian, Belarbi Rafik. Numerical analysis of hybrid ventilation performance depending on climate characteristics. International Journal of Ventilation 1: HybVent–Hybrid Ventilation Special Edition; 2003.
[76] Xu Xinhua, Wang Shengwei. Optimal simplified thermal models of building envelope based on frequency domain regression using genetic algorithm. Energy and Buildings 2007;39:525–36.
[77] Deng Kun, Barooah Prabir, Mehta Prashant G, Meyn Sean P. Building thermal model reduction via aggregation of states. In: American control conference (ACC), Marriott Waterfront, Baltimore, MD, USA; 2010. p. 5118–23.
[78] Bueno Bruno, Norford Leslie, Pigeon Grgoire, Britter Rex. A resistance–capacitance network model for the analysis of the interactions between the energy performance of buildings and the urban climate. Building and Environment 2012;54:116–25.
[79] Hazyuk Ion, Ghiaus Christian, Penhouet David. Optimal temperature control of intermittently heated buildings using model predictive control: Part I—building modeling. Building and Environment 2012;51:379–87.
[80] Axley James. Multizone airflow modeling in buildings: history and theory. HVAC and Research 2007; 13(6).
[81] Chen Qingyan. Ventilation performance prediction for buildings: a method overview and recent applications. Building and Environment 2009;44: 848–858.
[82] Kalogirou Soteris A. Use of TRNSYS for modelling and simulation of a hybrid PV-thermal solar system for cyprus. Renewable Energy 2001;23:247–60.
[83] Ibanez Manuel, Lazaro Ana, Zalba Belen, Cabeza Luisa F. An approach to the simulation of PCMs in building applications using TRNSYS. Applied Thermal Engineering 2005;25:1796–807.
[84] Zhai Zhiqiang (John), Johnson Mary-Hall, Krarti Moncef. Assessment of natural and hybrid ventilation models in whole-building energy simulations. Energy and Buildings 2011;43:2251–61.
[85] Goyal Siddharth, Barooah Prabir. A method for model-reduction of non-linear thermal dynamics of multi-zone buildings. Energy and Buildings 2012;47:332–40.
[86] MacDonald Iain Alexander. Quantifying the effects of uncertainty in building simulation. PhD thesis, Department of Mechanical Engineering University of Strathclyde; 2002.
[87] Parti Michael, Parti Cynthia. The total and appliance-specific conditional demand for electricity in the household sector. Bell Journal of Economics 1980;11:309–21.
[88] Lam Joseph C, Hui Sam CM, Chan Apple LS. Regression analysis of high-rise fully air-conditioned office buildings. Energy and Buildings 1997;26:189–97.
[89] Freire Roberto Z, Oliveira Gustavo HC, Mendes Nathan. Development of regression equations for predicting energy and hygrothermal performance of buildings. Energy and Buildings 2008;40:810–20.
[90] Lafrance G, Perron D. Evolution of residential electricity demand by end-use in quebec 1979–1989: a conditional demand analysis. Energy Studies Review 1994;6(Art 4).
[91] Tiedermann KH. Using conditional demand analysis to estimate residential energy use and energy savings. European Council for an Energy Efficient Economy Summer Study 2007;6:1279–83.
[92] Aranda Alfonso, Ferreira Germn, Mainar-Toledo MD, Scarpellini Sabina, Sastresa Eva Llera. Multiple regression models to predict the annual energy consumption in the Spanish banking sector. Energy and Buildings 2012;49:380–7.
[93] Givoni B, Vecchia F. Predicting thermal performance of occupied houses. In: PLEA 2001, Florianpolis, Brazil; 2001. p. 701–6.
[94] Krüger Eduardo, Givoni Baruch. Predicting thermal performance in occupied dwellings. Energy and Buildings 2004;36:301–7.
[95] Holland JH. Adaptation in natural and artificial systems. Ann Arbor; 1975.
[96] Caldas LG, Norford LK. Genetic algorithms for optimization of building envelopes and the design and control of HVAC systems. Journal of Solar Energy Engineering 2003;125:343–51.
[97] Magnier Laurent, Haghighat Fariborz. Multiobjective optimization of building design using TRNSYS simulations, genetic algorithm, and artificial neural network. Building and Environment 2010;45:739–46.
[98] Ooka Ryozo, Komamura Kazuhiko. Optimal design method for building energy systems using genetic algorithms. Building and Environment 2009;44:1538–44.
[99] Sadeghi Hossein, Zolfaghari Mahdi, Heydarizade Mohamad. Estimation of electricity demand in residential sector using genetic algorithm approach. International Journal of Industrial Engineering and Production Research 2011;22:43–50.
[100] Ozturk Harun Kemal, Ceylan Halim, Canyurt Olcay Ersel, Hepbasli Arif. Electricity estimation using genetic algorithm approach: a case study of Turkey. Energy 2005;30:1003–12.
[101] Datta D, Tassou SA, Marriott D. Application of neural networks for the prediction of the energy consumption in a supermarket. In: Proceedings of the Clima 2000 conference Brussels, Belgium; 1997.
[102] Y Lettvin J, Maturana HR, McCulloch WS, Pitts WH. What the frog's eye tells the frog's brain. Proceedings of the Institute of Radio Engineers 1959;47:1940–51.
[103] Kalogirou Soteris A, Bojic Milorad. Artificial neural networks for the prediction of the energy consumption of a passive solar building. Energy 2000;25:479–91.

[104] Kalogirou Soteris A, Neocleous Constantinos C, Schizas Christos N. Building heating load estimation using artificial neural networks. In: Proceedings of Clima 2000 Conference; 1997.
[105] Kalogirou Soteris A. Applications of artificial neural-networks for energy systems. Applied Energy 2000;67:17–35.
[106] Neto Alberto Hernandez, Fiorelli Flavio Augusto Sanzovo. Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. Energy and Buildings 2008;40:2169–76.
[107] Kwok Simon SK, Lee Eric WM. A study of the importance of occupancy to building cooling load in prediction by intelligent approach. Energy Conversion and Management 2011;52:2555–64.
[108] Escriva-Escriva Guillermo, Alvarez-Bel Carlos, Roldan-Blay Carlos, Alcazar-Ortega Manuel. New articial neural network prediction method for electrical consumption forecasting based on building end-uses. Energy and Buildings 2011;43:3112–9.
[109] Leung MC, Tse Norman CF, Lai LL, Chow TT. The use of occupancy space electrical power demand in building cooling load prediction. Energy and Buildings 2012;55:151–63.
[110] Cortes Corinna, Vapnik Vladimir. Support-vector networks. Machine Learning 1995;20:273–97.
[111] Dong Bing, Cao Cheng, Lee Siew Eang. Applying support vector machines to predict building energy consumption in tropical region. Energy and Buildings 2005;37:545–53.
[112] Lai Florence, Magoulès Frédéric, Lherminier Fred. Vapniks learning theory applied to energy consumption forecasts in residential buildings. International Journal of Computer Mathematics 2008;85:1563–88.
[113] 〈http://www.kxen.com/〉; 2007.
[114] Li Qiong, Meng Qinglin, Cai Jiejin, Yoshino Hiroshi, Mochida Akashi. Predicting hourly cooling load in the building: a comparison of support vector machine and different artificial neural networks. Energy Conversion and Management 2009;50:90–6.
[115] Li Qiong, Meng Qinglin, Cai Jiejin, Yoshino Hiroshi, Mochida Akashi. Applying support vector machine to predict hourly cooling load in the building. Applied Energy 2009;86:2249–56.
[116] Kavaklioglu Kadir. Modeling and prediction of turkeys electricity consumption using support vector regression. Applied Energy 2011;88:368–75.
[117] Paniagua-Tineo A, Salcedo-Sanz S, Casanova-Mateo C, Ortiz-Garcia EG, Cony MA, Hernandez-Martn E. Prediction of daily maximum temperature using a support vector regression algorithm. Renewable Energy 2011;36:3054–60.
[118] Che Jinxing, Wang Jianzhou, Wang Guangfu. An adaptive fuzzy combination model based on self-organizing map and support vector regression for electric load forecasting. Energy 2012;37:657–64.
[119] Chen Ji-Long, Liu Hong-Bin, Wu Wei, Xie De-Ti. Estimation of monthly solar radiation from measured temperatures using support vector machines—a case study. Renewable Energy 2011;36:413–20.
[120] Saha Bhaskar, Goebel Kai, Poll Scott, Christophersen Jon. An integrated approach to battery health monitoring using Bayesian regression and state estimation. Autotestcon, IEEE; 2007. p. 646–53.
[121] Teeter Jason, Chow Mo-Yuen. Application of functional link neural network to HVAC thermal dynamic system identification. IEEE Transactions on Industrial Electronics 1998;45(1).
[122] Paris Benjamin, Eynard Julien, Grieu Stphane, Polit Monique. Hybrid PID-fuzzy control scheme for managing energy resources in buildings. Applied Soft Computing 2011;8:5068–80.
[123] Nassif Nabil, Kajl Stanislaw, Sabourin Robert. Optimization of HVAC control system strategy using two-objective genetic algorithm. HVAC & R Research 2005;11:459–86.
[124] Lauret Philippe, Boyer Harry, Riviere Carine, Bastide Alain. A genetic algorithm applied to the validation of building thermal models. Energy and Buildings 2005;37:858–66.
[125] Boyer H, Garde F, Gatina JC, Brau J. A multimodel approach to building thermal simulation for design and research purposes. Energy and Buildings 1998;28:71–8.
[126] Miranville Frederick Winkelmannic, Boyer Harry, Mara Thierry, Garde Franois. On the thermal behavior of roof-mounted radiant barriers under tropical and humid climatic conditions: modelling and empirical validation. Energy and Buildings 2003;35:997–1008.
[127] Znouda Essia, Ghrab-Morcos Nadia, Hadj-Alouane Atidel. Optimization of mediterranean building design using genetic algorithms. Energy and Buildings 2007;39:148–53.
[128] Ghrab-Morcos N. CHEOPS: a simplified tool for thermal assessment of Mediterranean residential buildings in hot and cold seasons. Energy and Buildings 2005;37:651–62.
[129] Wang Shengwei, Xu Xinhua. Simplified building model for transient thermal performance estimation using GA-based parameter identification. International Journal of Thermal Sciences 2006;45:419–32.
[130] Wang Shengwei, Xu Xinhua. Parameter estimation of internal thermal mass of building dynamic models using genetic algorithm. Energy Conversion and Management 2006;47:1927–41.
[131] Tuhus-Dubrow Daniel, Krarti Moncef. Genetic-algorithm based approach to optimize building envelope design for residential buildings. Building and Environment 2010;45:1578–81.
[132] Winkelmann FC, Birsdall BE, Bull WF, Ellington KL, Erdem AE, Hirsh JJ. DOE-2 supplement, version 2.1e, Technical Report lbl-34947. Lawrence Berkeley National Laboratory, Berkeley, CA; 1993.
[133] Siddharth V, Ramakrishna PV, Geetha T, Sivasubramaniam Anand. Automatic generation of energy conservation measures in buildings using genetic algorithms. Energy and Buildings 2011;43:2718–26.
[134] Sahu M, Bhattacharjee B, Kaushik SC. Thermal design of air-conditioned building for tropical climate using admittance method and genetic algorithm. Energy and Buildings 2012;53:1–6.
[135] Yang Zhenyu, Li Xiaoli, Bowers Chris P, Schnier Thorsten, Tang Ke, Yao Xin. An efficient evolutionary approach to parameter identification in a building thermal model. IEEE Transactions on Systems 2012;42:957–69.
[136] deWit M. HAMBase: heat, air and moisture model for building and systems evaluation. Eindhoven, The Netherlands: The Eindhoven University Press; 2006.
[137] Lam Joseph C, Hui Sam CM. Sensitivity analysis of energy performance of office buildings. Building and Environment 1996;31:27–39.
[138] Mendes Nathan, Oliveira Ricardo CLF, Santos Gerson Henrique Dos. Energy efficiency and thermal comfort analysis using the PowerDomus hygrothermal simulation tool. In: Building simulation IBPSA conference; 2005.
[139] Xu Xiaoqi, Taylor John E, Pisello Anna Laura, Culligan Patricia J. The impact of place-based affiliation networks on energy conservation: an holistic model that integrates the influence of buildings, residents and the neighborhood context. Energy and Buildings 2012;55:637–46.
[140] Lee JW, Jung HJ, Park JY, Lee JB, Yoon Y. Optimization of building window system in asian regions by analyzing solar heat gain and daylighting elements. Renewable Energy 2013;50:522–31.
[141] Madu Christian N. Simulation in manufacturing: a regression metamodel approach. Computers & Industrial Engineering 1990;18:381–9.
[142] Jin Ruichen, Du Xiaoping, Chen Wei. The use of metamodeling techniques for optimization under uncertainty. Structural and Multidisciplinary Optimization 2003;25:99–116.
[143] Villanueva D, Picard G, Le Riche R, Haftka RT. Optimisation multi-agent par partitionnement adaptatif de lespace de conception. 20èmes Journes Francophones sur les Systmes Multi-Agents; 2012.
[144] Clarke JA, Cockroft J, Conner S, Hand JW, Kelly NJ, Moore R, et al. Simulation-assisted control in building energy management systems. Energy and Buildings 2002;34:933–40.
[145] Trocmé Maxime. Aide au choix de conception de bâtiments économes en énergie. PhD thesis, Ecole Nationale Supérieure des Mines de Paris; 2009.